The impact of single sex Schools on Female wages at age 33

```
. regress lhw_5 ssex_3 payrsed mayrsed if male==0 & touse_2==1;

     Source |       SS       df       MS                  Number of obs =    1324
-------------+------------------------------              F(  3,  1320) =   27.29
       Model |  18.5473889     3  6.18246296              Prob > F      =  0.0000
   Residual |  299.030192  1320  .226538024               R-squared     =  0.0584
-------------+------------------------------              Adj R-squared =  0.0563
       Total |   317.57758  1323  .240043523              Root MSE      =  .47596

------------------------------------------------------------------------------
      lhw_5 |                                Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    Single Sex School           ssex_3 |   .2480322   .0290271     8.54   0.000     .1910879    .3049765
Father's Years of Education    payrsed |   .002115    .0054614     0.39   0.699    -.008599     .012829
Mother's Years of Education    mayrsed |   .0061507   .0057607     1.07   0.286    -.0051505    .0174518
Constant                         _cons |  1.548435    .0314873    49.18   0.000    1.486664    1.610206
------------------------------------------------------------------------------
```

The impact of single sex Schools on Female wages at age 33. Including an ability indicator

```
.  regress lhw_5 ssex_3 lowabil  payrsed mayrsed if male==0 & touse_2==1;

     Source |       SS       df       MS              Number of obs =    1324
-------------+------------------------------           F(  4,  1319) =   37.50
      Model |  32.4283351     4  8.10708377           Prob > F      =  0.0000
   Residual |  285.149245  1319  .216185933           R-squared     =  0.1021
-------------+------------------------------           Adj R-squared =  0.0994
      Total |   317.57758  1323  .240043523           Root MSE      =  .46496

------------------------------------------------------------------------------
      lhw_5 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

		Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Single Sex School	**ssex_3**	**.2126029**	.0286988	7.41	0.000	.1563027	.2689032
Low Ability	lowabil	-.2159654	.0269518	-8.01	0.000	-.2688385	-.1630922
Father’s Years of Education	payrsed	.0013796	.005336	0.26	0.796	-.0090883	.0118475
Mother’s Years of Education	mayrsed	.0053448	.0056284	0.95	0.342	-.0056969	.0163865
Constant	_cons	1.649788	.0332586	49.60	0.000	1.584543	1.715034

The impact of single sex Schools on Female wages at age 33. Including an ability indicator and School type

Source	SS	df	MS		
Model	38.4255269	5	7.68510537	Number of obs =	1324
Residual	279.152054	1318	.211799737	F(5, 1318) =	36.28
Total	317.57758	1323	.240043523	Prob > F =	0.0000
				R-squared =	0.1210
				Adj R-squared =	0.1177
				Root MSE =	.46022

	lhw_5	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Single Sex School	**ssex_3**	**.1288113**	.0324787	3.97	0.000	.0650957	.192527
Low Ability	lowabil	-.1793011	.0275525	-6.51	0.000	-.2333525	-.1252496
Selective School	selecsch	.1979693	.0372037	5.32	0.000	.1249843	.2709543
Father’s Years of Education	payrsed	.0003462	.0052851	0.07	0.948	-.010022	.0107143
Mother’s Years of Education	mayrsed	.0050315	.0055714	0.90	0.367	-.0058982	.0159612
Constant	_cons	1.629437	.0331409	49.17	0.000	1.564422	1.694451

Correlations Between the Various Variables

	ssex_3	lowabil	selecsch
ssex_3	1.0000		
lowabil	-0.1618	1.0000	
selecsch	0.5007	-0.2927	1.0000

Stepwise Regression

•Consider again the regression of a dependent variable (*Y)* on two regressors (X_1 and X_2).

$$Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + u_i$$

•The OLS estimator of one of the coefficients b_1 can be written as a function of sample variances and covariances as

$$\hat{b}_1 = \frac{\mathrm{cov}(Y, X_1) Var(X_2) - \mathrm{cov}(X_1 X_2)\,\mathrm{cov}(Y, X_2)}{Var(X_1) Var(X_2) - \mathrm{cov}(X_1 X_2)^2}$$

- We can divide all terms by $Var(X_2)$. This gives

$$\hat{b}_1 = \frac{\text{cov}(Y, X_1) - \dfrac{\text{cov}(X_1 X_2)}{Var(X_2)}\text{cov}(Y, X_2)}{Var(X_1) - \dfrac{\text{cov}(X_1 X_2)}{Var(X_2)}\text{cov}(X_1 X_2)}$$

- Now notice that $\dfrac{\text{cov}(X_1 X_2)}{Var(X_2)}$

is the regression coefficient one would obtain if one were to regress X_1 on X_2 i.e. the OLS estimator of b_{12} in

$$X_{i1} = a_{12} + b_{12} X_{i2} + v_i$$

We call this an "auxiliary regression"

- So we substitute this notation in the formula for the OLS estimator of b_1 to obtain

$$\hat{b}_1 = \frac{\mathrm{cov}(Y, X_1) - \hat{b}_{12}\,\mathrm{cov}(Y, X_2)}{Var(X_1) - \hat{b}_{12}\,\mathrm{cov}(X_1 X_2)}$$

- Note that if the coefficient $\hat{b}_{12}$ in the "auxiliary regression" is zero we are back to the results from the simple two variable regression model

- Lets look at this formula again by going back to the summation notation. Canceling out the Ns we get that:

$$\hat{b}_1 = \frac{\sum_{i=1}^{N}(X_{i1} - \overline{X}_1)(Y_i - \overline{Y}) - \hat{b}_{12}\sum_{i=1}^{N}(X_{i2} - \overline{X}_2)(Y_i - \overline{Y})}{\sum_{i=1}^{N}(X_{i1} - \overline{X}_1)^2 - \hat{b}_{12}\sum_{i=1}^{N}(X_{i1} - \overline{X}_1)(X_{i2} - \overline{X}_2)} =$$

$$\frac{\sum_{i=1}^{N}\left[(X_{i1} - \overline{X}_1) - \hat{b}_{12}(X_{i2} - \overline{X}_2)\right](Y_i - \overline{Y})}{\sum_{i=1}^{N}\left[(X_{i1} - \overline{X}_1) - \hat{b}_{12}(X_{i2} - \overline{X}_2)\right]^2}$$

- To derive this result you need to note the following

$$\hat{b}_{12}^2 \sum (X_{i2} - \overline{X}_2)^2 = \hat{b}_{12} \frac{\sum (X_{i2} - \overline{X}_2)(X_{i1} - \overline{X}_1)}{\sum (X_{i2} - \overline{X}_2)^2} \sum (X_{i2} - \overline{X}_2)^2 =$$

$$\hat{b}_{12} \sum (X_{i2} - \overline{X}_2)(X_{i1} - \overline{X}_1)$$

- This implies that

$$\sum_{i=1}^{N} \left[(X_{i1} - \overline{X}_1) - \hat{b}_{12}(X_{i2} - \overline{X}_2) \right]^2 = \sum_{i=1}^{N} (X_{i1} - \overline{X}_1)^2 - \hat{b}_{12} \sum_{i=1}^{N} (X_{i1} - \overline{X}_1)(X_{i2} - \overline{X}_2)$$

- Now the point of all these derivations can be seen if we note that

$$\hat{v}_i = (X_{i1} - \bar{X}_1) - \hat{b}_{12}(X_{i2} - \bar{X}_2)$$

Is the residual from the regression of X_1 on X_2 .

- This implies that the OLS estimator for b_1 can be obtained in the following two steps:
  - Regress X_1 on X_2 and obtain the residuals from this regression
  - Regress Y on on these residuals
- Thus the second step regression is

$$Y_i = b_1 \hat{v}_i + u_i$$

- This procedure will give identical estimates for $\hat{b}_1$ as the original formula we derived.

- The usefulness of this stepwise procedure lies in the insights it can give us rather than in the computational procedure it suggests.

What can we learn form this?

- Our ability to measure the impact of X_1 on Y depends on the extent to which X_1 varies over and above the part of X_1 that can be "explained" by X_2 .

- Suppose we include X_2 in the regression in a case where X_2 does not belong in the regression, i.e. in the case where b_2 is zero. This approach shows the extent to which this will lead to **efficiency loss** in the estimation of b_1.

- **Efficiency Loss** means that the estimation precision of $\hat{b}_1$ declines as a result of including X_2 when X_2 does not belong in the regression.

The efficiency loss of including irrelevant regressors

- We now show this result explicitly.
- Suppose that b_2 is zero.
- Then we know by applying the Gauss Markov theorem that the efficient estimator of is b_1 is

$$\tilde{b}_1 = \frac{\operatorname{cov}(X_1 Y)}{Var(X_1)}$$

- Instead, by including X_2 in the regression we estimate b_1 as

$$\hat{b}_1 = \frac{\sum_{i=1}^{N}\left[(X_{i1} - \overline{X}_1) - \hat{b}_{12}(X_{i2} - \overline{X}_2)\right](Y_i - \overline{Y})}{\sum_{i=1}^{N}\left[(X_{i1} - \overline{X}_1) - \hat{b}_{12}(X_{i2} - \overline{X}_2)\right]^2}$$

- The Gauss Markov theorem directly implies that $\tilde{b}_1$ cannot be less efficient that $\hat{b}_1$

- However we can show this directly.
- We know that the variance of $\tilde{b}_1$ is

$$Var(\tilde{b}_1) = \frac{s^2}{NVar(X_1)}$$

- By applying the same logic to the 2nd step regression we get that

$$Var(\tilde{b}_1) = \frac{s^2}{NVar(\hat{v})}$$

- Since $\hat{v}$ is the residual from the regression of X_1 on X_2 it must be the case that the variance of $\hat{v}$ is no larger than the variance of X_1 itself.
- Hence $Var(X_1) \geq Var(\hat{v})$
- This implies $Var(\tilde{b}_1) \leq Var(\hat{b}_1)$
- Thus we can state the following result:
- <u>Including an unnecessary regressor, which is correlated with the others, reduces the efficiency of estimation of the the coefficients on the other included regressors.</u>

Summary of results

- Omitting a regressor **which has an impact** on the dependent variable and is correlated with the included regressors leads to "**omitted variable bias"**

- Including a regressor **which has no impact** on the dependent variable and is correlated with the included regressors leads to a **reduction in the efficiency** of estimation of the variables included in the regression.